(54) Title: SYSTEM AND METHOD FOR RETRIEVING INFORMATION FROM A PLURALITY OF DATABASES

(57) Abstract

This invention enables a user (5) to obtain information from a large number of commercial databases (20-24). In practicing the invention, the user (5) selects an area of interest (40) and enters a search request (42). The search request (42) includes at least one word for which the user (5) desires to search. In one embodiment of the invention, the system selects a set of at least two databases (two of 20-24), automatically executes the search request (46) in each database, and presents the results to the user (5, 48). In another aspect of the invention, the user (5) selects a database directly, and employs a set of standardized commands (see figure 4) for any database selected. The system translates these standardized commands into the equivalent commands recognized by each database, without the intervention or knowledge of the user (5). The user (5) can thus communicate with a variety of databases (20-24) using the same command set (see figure 4). In another embodiment, the invention guides the user (4) in reformulating a search which retrieved either no documents or too many documents. The invention also includes a method of determining which of the retrieved documents are likely to be the most relevant.

1

SYSTEM AND METHOD FOR RETRIEVING INFORMATION
FROM A PLURALITY OF DATABASES

CROSS-REFERENCE TO PRIOR APPLICATIONS

This is a continuation-in-part of U.S. Patent Application Serial No. 231,055, filed August 11, 1988, entitled "System for Retrieving Information From a Plurality of Remote Databases", which is a continuation of Serial No. 664,167, filed Oct. 24, 1984, now U.S. Patent No. 4,774,655.

BACKGROUND OF THE INVENTION

This invention relates to the field of information retrieval, and especially the retrieval of information from one or more databases.

The field of information retrieval has advanced to the point where anyone having a personal computer, a modem, and access to a telephone line can obtain information on virtually any topic, from thousands of commercial databases, without leaving home. Many owners of large compilations of information have provided their information in the form of computer databases, and these databases can be interrogated, for a fee, by a remote user.

Groups of related and unrelated databases are sometimes arranged in "families". A single vendor, or "host", may provide access to a family of databases, all of which can be reached through the same telephone number, and with the same identification number and password. For example, Dialog Information Services, Inc., of California, provides access to a large family of databases, dealing with many different subjects, under the service mark DIALOG. In this specification, the term "family" means a set of related or unrelated databases available on one particular host. As used herein, the terms "family" and "host" are virtually interchangeable.

To use each database, one must know, in general, a telephone number, an identification number, and (in most cases) a password. One must also know the "language" of the database being searched. As applied to database searching, the term "language" includes at least three aspects. First, a database language includes the particular syntax applicable to search requests in that database. A search request includes a word or group of words, connected by various logical (typically Boolean) operators (e.g. AND, OR, NOT, etc.). The user transmits a search request to a database in order to retrieve all items in the database which contain the specified logical combination of terms of the search request. In general, different database families have different rules of search syntax.

Secondly, a database language includes a command set. As used in this specification, the term "command set" means a group of commands used to conduct searches in a database. A command is needed to direct the performance of each aspect of database searching. Thus, for example, a command is needed to tell the system to begin the search, and another command is needed to direct the system to display the results of a search.

Thirdly, a database language includes the field structure of the database. Each database is arranged by fields, i.e. searchable segments of documents. For example, one might want to search for articles in a database by searching according to authors or titles or abstracts. The "author", "title", and "abstract" segments of a document are called "fields". In general, each database provides a different set of fields which can be searched, and a different set of "field tags" which identify particular fields.

In theory, each database could have a different syntax, a different

command set, and a different field structure, since each database is, in general, created by a different entity. Thus, to use all of the hundreds or thousands of available databases, it would be necessary to learn all of the rules for syntax, command sets, and field structures for each one.

The problem of nonuniformity of database languages was ameliorated somewhat by the creation of database families, such as DIALOG, mentioned above. Within a particular family, the syntax and command sets are generally the same for each database. As long as a user wishes to search only within one database family, the user needs to know only one search syntax and one command set. Of course, the user would also need a database catalog, in order to know what databases are available.

However, even within a given database family, the field structures of the various databases are, in general, very different. Moreover, if one desires to search in other families of databases, or in single databases not belonging to a family, one must, in general, learn a new search syntax, a new command set, and a new field structure. It is also necessary to establish a separate account with each different database or database family. And, of course, it is necessary for the user to know, in advance, what databases are available and what they contain.

Database searching presents additional problems not discussed above. Frequently, a search yields no documents, or "postings". The null result could be due to misspelled words in the search request, or to other factors. Just as frustrating is the case where a search produces too many postings. It is tedious and expensive to browse through one hundred or more documents retrieved from a database search. Even if the number of postings is not unreasonably large, it is still necessary to browse through them individually to determine which documents are most relevant. The fees incurred for viewing documents can often exceed the costs of

4

performing the search, because most databases or hosts charge a fixed fee
for each document displayed to the user. Search systems of the prior art
have not provided an efficient way of determining, in advance, which re-
trieved documents are likely to be the most relevant. The present inven-
tion is therefore also directed to the solution of the above-mentioned
problems.

U.S. Patent No. 4,774,655, the disclosure of which is incorporated
by reference herein, addresses the problems of nonuniformity among data-
bases and database families. The system described in that patent is an
intermediary between the user and the databases being searched. The sys-
tem described in the patent automatically selects a database in which to
search, and then performs the search. The user does not make direct con-
tact with the database, and does not need to know the particular syntax
of the database, or even its identity. The system accepts, from the
user, a search request, written in a simple and standardized format, and
automatically translates the search request into the syntax appropriate
to the database in which the system has chosen to search. The system
minimizes the time that it is directly connected ("on line") to the data-
base, and presents the search results to the user after the connection
with the database has been terminated. In an alternative embodiment,
also described in the patent, the user selects a database directly, but
still uses the system as an intermediary, thereby avoiding the need to
establish a separate account with each database. It is also possible to
take advantage of the ability of the system to translate a search request
into the syntax of a selected database or database family, while still
maintaining full control over the identity of the database or family se-
lected.

The present specification discloses enhancements to the basic system described in the cited patent. One of these enhancements allows a user to search through a large number of databases in one session. Another embodiment improves on the standardization described in the cited patent, by enabling the user to search through many different selected databases using a standardized set of commands. Other features reduce the likelihood of null search results, and also deal with the problem of too many retrieved documents. Another feature provides an economical method of determining, in advance, which retrieved documents are likely to be the most relevant. With these features, the system is even more convenient to use than systems of the prior art, and database searches are more likely to yield meaningful results.

SUMMARY OF THE INVENTION

The present invention enables a user to obtain information from one or more databases. The user communicates with a central computer, programmed according to the invention, and the computer establishes direct communication with the databases. In general, the user is not located at the same place as the central computer, and the central computer is not at the same location as the various databases. However, the invention is not limited by the distances between the user, the central computer, and the databases.

In one aspect of the invention, the user first transmits an area of interest and a search request to the central computer. The area of interest can be selected from a menu of available subject areas. The search request includes one or more words, connected by logical operators, such as Boolean operators or proximity connectors. The object of

6

the search is to locate all documents which contain the indicated logical combination of words of the search request. The central computer executes the search request in each of a set of databases associated with the user's area of interest. The databases in this set are predetermined, and are stored in the memory of the central computer, one set being associated with each area of interest. The databases may be located on different hosts, or they may reside on the same host. The central computer establishes communication separately with each database, translates the search request into the syntax of that database, if necessary, and executes the search. Normally, the searches are done sequentially, in each database, but it is also possible to do the searches simultaneously if the central computer is capable of establishing multiple simultaneous connections. After completing the searches, the system displays, to the user, the number of items retrieved from each of the databases, and gives the user the opportunity to view these items in detail. When the user wishes to view a specific retrieved item, the system may simply re-connect to the indicated database, search for the indicated item, and download the item for display to the user. It is also possible to allow the user to browse through the retrieved item while the central computer is still connected to the database.

In another aspect of the invention, the system provides a standardized set of commands for searching in databases. As used herein, "commands" mean the instructions given by the user to guide the operation of a search. Commands are used to generate searches, display results, review the sets of documents retrieved, terminate the search session, etc. In this embodiment, the system translates the standardized commands entered by the user into the commands appropriate to the specific database

selected by the user. Thus, although the user may want to select a par-
ticular database for searching, the user need not learn the particular
command set for that database. The central computer establishes the con-
nection with the selected database, and the user does not directly deal
with that database.

The latter embodiment may be combined with other embodiments. Thus,
the standardized command set can be used even in the case where the sys-
tem chooses the database for the user. Also, the user may select a data-
base but leave the translation of both the search request and the com-
mands to the system. In the latter case, the user need not know anything
about the database other than its name.

In another embodiment of the invention, the system guides the user
in reformulating a search, in those cases where a search produces no
postings or too many postings. The system identifies words of the search
request which were not found in any document, and gives the user the op-
portunity to modify or delete those words. In cases where each word of
the search request yielded postings, but the overall request did not
yield any postings, the system provides suggestions to the user for
broadening the search. In the case of too many postings, the system pro-
vides the user with suggestions on how to narrow the search, such as by
imposing more stringent field or proximity restrictions on portions of
the search request.

The invention also includes a system and method for determining
which retrieved documents are likely to be the most relevant. This
method is performed without browsing through the actual documents. In-
stead, the system considers, for example, the fields in which the search
terms were found. Thus, if a search term is found in the title of the
document, it is likely that the document is more relevant than if the

8

term appeared only in an abstract.

It is therefore an object of the invention to provide a system and method for retrieving information from databases, wherein a user's search request is automatically executed in a group of databases selected by the system.

It is another object to provide a method and system which chooses a group of databases which are relevant to an area of interest selected by a user.

It is another object to provide a method and system wherein a search can be executed in several databases simultaneously.

It is another object to provide a method and system which includes automatic selection of a host, where the desired database is found on more than one host.

It is another object to enable a user to communicate with a plurality of commercial databases using a standardized set of commands.

It is another object to provide a system wherein the standardized set of commands includes a standardized set of field "tags".

It is another object to provide a system wherein the user can choose a particular database, but wherein the user does not need to learn the search syntax or command set for that database, and wherein the user does not need to establish a separate account with that database.

It is another object to provide a method and system for alleviating the problems associated with too few or too many documents retrieved in a database search.

It is another object to provide an economical method of evaluating which documents, retrieved from a search, are likely to be the most relevant.

9

It is another object to increase the ease and effectiveness of database searching.

Other objects and advantages of the invention will be apparent to those skilled in the art from a reading of the following brief description of the drawings, the detailed description of the invention, and the appended claims.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a block diagram showing a possible configuration of the system of the present invention.

Figure 2 is a flow chart showing the operation of the embodiment of the invention wherein the system executes a search automatically in each of several databases.

Figure 3 is a block diagram showing an alternative configuration of the system, wherein the central computer can perform searches simultaneously in more than one database.

Figure 4 is a chart containing a hypothetical standardized set of commands, used for searching in databases, and showing the meaning of each command.

Figure 5 is a flow chart illustrating an embodiment of the invention which assists the user in reformulating a search that has retrieved too few or too many documents.

10

DETAILED DESCRIPTION OF THE INVENTION

Figure 1 is a block diagram showing one possible configuration of the system of the present invention. The system may be operated from terminal 5, having modem 7. Terminal 5 may be simply a personal computer, or it may be a so-called dedicated terminal (or "dumb terminal"), whose sole capability is to transmit and receive information. Modem 7, which is of conventional design, and which is assumed to include a telephone and dialing mechanism, converts the digital signals from the terminal 5 into analog signals suitable for transmission over telephone line 9. Modem 7 also converts incoming analog signals from line 9 into signals having digital format, which can then be interpreted by terminal 5.

When modem 7 dials the telephone number of the system, the call is answered by switching unit 11. Switching unit 11 allocates the call to one of a plurality of identically-programmed computers 13, each computer having an incoming modem 15. Switching unit 11 is of standard design, and is pre-programmed to indicate to the caller if all available computers are currently busy. In the preferred embodiment, computers 13 comprise identical central processing units (CPUs) which may be conveniently arranged in the same housing. Other arrangements are possible, however.

Each computer is connected to an outgoing modem 17. Outgoing modems 17 are also of standard design, and are equipped with automatic telephone dialing mechanisms. Through modems 17, the computers 13 can establish communication, via telephone lines 19, with one of many remote commercial databases, illustrated symbolically as blocks 20, 21, 22, 23, and 24.

The modems shown in Figure 1 can be replaced by equivalents. Thus, it is possible to provide direct digital communication between computers 13 and the various databases, by the use of alternative technologies

SUBSTITUTE SHEET

11

which are available, such as packet assemblers and disassemblers ("pads"). Also, the connection between terminal 5 and computer 13 may include a conventional analog connection between the terminal and a given node in a telecommunications network, and a digital connection between that node and one of computers 13. Such alternative arrangements are within the scope of the present invention.

Each of the databases 20-24 represents, in general, a separate and independent computer system at some remote location. The information stored in each database may be kept on disks or other storage media, and the searching through the respective databases is governed by the particular computer system for that database. The system of the present invention assumes the existence of such multiple databases, and operates with any or all of them, regardless of the specific configuration of the computer system of each database. The system communicates with these databases in the same manner as would any other user of such databases.

Connected in parallel to all the computers 13 is master CPU 30. Master CPU 30 is a computer which controls the overall operation of the system. The master CPU continually checks that all computers 13 are operating properly. It also can be programmed, in conjunction with switching unit 11, to guide an incoming user call to the appropriate computer. The master CPU 30 also serves to direct the various computers to retrieve information from disk 31, or to print billing information on printer 32.

The components shown between dotted lines 1 and 2 indicate the components which are "local", i.e. grouped at a central location. The components shown outside the dotted lines are "remote", i.e. located elsewhere. Normally, terminal 5 will be located at a substantial distance from the system, possibly thousands of miles away. Databases 20-24 are

12

also, in general, found in computers located in other places. However, it should be understood that it is possible that one or more of the "remote" components could be physically located near the central location, without changing the manner of operation of the system. Thus, it is possible to operate the system from a terminal located near computers 13. It is also possible that one or more of databases 20-24 could be located in the same general area. The invention will operate regardless of the physical locations of the remote terminal and the databases. Also, the number of CPUs can be varied. The invention will work with even one central CPU, having a modem connected to the line coming from the user, and another modem connected to a line connected to a database. The databases 20-24 can be individual databases, or they can be database families. In the latter case, the system determines which database to search, within a particular family.

U.S. Patent No. 4,774,655 describes several methods of using the system illustrated in Figure 1. In one such method, computers 13 are programmed to display, to the user, one or more menus from which the user selects an area of interest. The user selects one item from each such menu, whereupon the system automatically chooses one database for searching. The choice of a database is made from a stored "decision tree" which associates one database with every possible combination of menu selections by the user. The user then enters a search request, which comprises one or more words, connected by logical operators. The aim of the search is to find documents which contain the indicated logical combination of words in the search request. The words of the search request can also be connected by non-Boolean operators, and all such alternatives should be considered equivalents for purposes of this invention.

The system described in the cited patent then automatically dials the telephone number of the selected database, and establishes a connection, using an identification number and password which has been stored in memory. Note that it is the system, not the ultimate user, which is the customer of the database. The system automatically translates the search request into the search syntax of the selected database, and transmits the search request to the database. The system may then download, to its own memory, some or all of the results of the search, and terminates the connection with the database. The system then displays the downloaded results of the search to the user. This displaying step may include allowing the user to browse electronically through one or more articles or other documents retrieved in the search. This browsing is therefore done after the system is disconnected ("off-line") from the database. The system can also print a bill, using printer 32, based on credit card information previously supplied by the user. Note that the user need establish only one account, i.e. a credit card account, to gain access to a wide variety of databases. The individual databases do not "see" the user as a customer.

Further variations to the above-described process are explained in the cited patent. In general, each such variation uses substantially the same general arrangement of hardware, the difference being in the programming of the computers.

Some databases are included in more than one database family, i.e. they are present on more than one host. The database selection step may therefore also include the step of choosing a host on which to search.

The choice of a host can be made in the following manner. For each database, the system stores a list of all hosts on which that database is available. The hosts on each list are ranked in a predetermined order

which is based on considerations of economics and efficiency. When the system decides to search a particular database, it will attempt to gain access to that database through the first host on the list. If the first choice host is busy or otherwise unavailable, the system then tries to connect to the second host on the list, and so on until a connection to a host is made.

The rankings of hosts may vary depending on the time of the day or the day of the week, or on other factors. For example, the host which is least expensive on a weekday may not be the least expensive on a weekend. Thus, the system can be programmed with several lists of hosts, for each database, each list being appropriate to a different time or day. Before connecting to a database, the system checks the time of day and/or day of the week, and then refers to the ranking of hosts which is appropriate to the day or time. Thus, the system is not limited to a single ranking of hosts for each database.

The criteria for ranking hosts can vary, and need not be limited to considerations of economics. It may be that, for a given database, a particular host is most efficient, despite the fact that it is not the least expensive. Also, the hosts can be chosen by means other than from predetermined lists or rankings.

Various factors can cause the system to connect to a host other than the first choice on the list. As mentioned above, the first choice host may be temporarily busy or unavailable. Also, the data telecommunications network, or "common carrier", which links the system to a database or host, may be unavailable. Also, the system may have exhausted its supply of available passwords for a particular host, due to a large number of users doing searches on databases in that host, and it may be im-

possible to obtain further access to the host for this reason.   Of

course, the user is unaware of which host the system chooses.

Figure 2 is a flow chart illustrating the operation of an embodiment

of the present invention which is a modification of the procedures de-

scribed above and in the cited patent. It is assumed, in Figure 2, that

the user has already established connection with the system, and has

transmitted satisfactory credit card information, or other identifica-

tion. Then, in block 40, the user begins by selecting an area of inter-

est from a menu displayed by the system. In block 42, the user enters a

search request. The functions of blocks 40 and 42 could be performed in

reverse order, if desired. In block 44, the system chooses a set of

databases in which searches will be made. The system stores, in its mem-

ory, a data file comprising a list of areas of interest and databases.

With each area of interest, the system associates a set of databases, all

of which contain information relating to that area. This set of databas-

es is fixed for each possible selection of area of interest by the user.

Thus, as in the cited patent, the user need not know, in advance, what

databases to search, and the selection of databases is entirely auto-

matic.

In block 46, the system then searches through each database in the

selected set. Each search may be performed in substantially the same

manner as described for an individual search in the cited patent. That

is, the system automatically translates the search request to conform to

the syntax of the database being searched, establishes connection with

the database (transmitting an appropriate identification number and pass-

word, if necessary), and transmits the translated search request to the

database. If a particular database is a member of a family, the system

gains access to the family, choosing that family according to the method

described above, and then transmits the correct database identifier, so as to connect to the desired database.

In the embodiment of Figure 2, however, the system does not immediately present retrieved documents to the user. Instead, the system temporarily stores the number of documents, or other items, retrieved from the database and disconnects from the database. The process is repeated for all of the databases in the set to be searched.

After searching through the entire set of selected databases, the system displays, to the user, in block 48, a summary showing the name of each database that has been searched and the number of items retrieved from each database. The system then asks, in test 50, if the user wants to browse through any or all of the retrieved items. If the answer is no, then the program will stop, as shown in block 52. If the answer is yes, the system accepts a choice, from the user, in block 54, indicating which item(s) from which database should be viewed.

The user's choice, made in block 54, can include a direction to download all documents retrieved from one of the databases. However, this alternative can be very expensive if the number of documents is large, as the database charge is based, in part, on the number of items downloaded. Thus, the system is preferably programmed to allow the user to specify which documents from each database should be displayed. For example, the user may specify the desire to view "document numbers 3-6", or some other subset, retrieved from "database number 6". Many other equivalent retrieval schemes can also be used.

After receiving a choice from the user, the system reconnects to the selected database, in block 56. As before, the system translates and executes the search again, in block 58, but this time the system down-

loads the selected items to its memory, as indicated in block 60, and

terminates the connection with the database. In block 62, the system

displays the retrieved documents to the user. The function of block 62

may include an interactive display, allowing the user to browse electron-

ically through the document(s). During this displaying step, the user

remains connected to the system but the system is not connected to any

database, so no additional database charges are incurred. When the user

is finished viewing the item, the program can return to test 50, and the

document-viewing process can be repeated.

In the procedure discussed above, the system searches a set of data-

bases one at a time. It is also possible to search through many databas-

es virtually simultaneously. However, in the latter case, it would be

necessary for each computers 13 to be connected, through separate modems,

to a plurality of databases. Alternatively, computers 13 can be arranged

as one larger computer. In either case, the computer would be subdivided

into separate processors, or otherwise programmed on a time-sharing

basis, so that signals could be passed back and forth between the system

and each of several databases virtually simultaneously. The computer

could also remain connected to all of the databases continuously, espe-

cially if there is a constant stream of search requests covering all or

most of the available databases or database families.

Figure 3 illustrates one possible arrangement for the latter embodi-

ment. Terminal 70 is connected, by a telephone line, to CPU 72, which

can establish connections with databases such as 78, 80, and 82. Dotted

lines 74 and 76 indicate the boundaries between the system and the compo-

nents external to the system. CPU 72 can send data through any of modems

84, 86, 88, 90, and 92. The CPU automatically connects to as many modems

as it needs to do the simultaneous searches. The remaining modems can be

18

used for the searches of another user. Clearly the number of modems can be varied; if a single large CPU is used for all users, the number of modems will be quite large, in general.

If, in the embodiment of Figure 3, the system needs to search through two or more databases on the same host, the searches can be done sequentially, using the modem which is connected to that host. Alternatively, the CPU can appropriate another modem, so that two modems may be connected to the same host (but not the same database) at the same time.

Dotted lines 94 and 96 symbolically illustrate the variability of the number of modems being employed for a given user at any one time. In the example of Figure 3, the CPU is searching only three databases, and modems 90 and 92 are not being used. But one or both of these modems could have been used if it had been necessary to search four or five databases. What is important is that the CPU be programmed to take command of the number of modems necessary to establish the desired simultaneous connections with databases and/or hosts.

The selection of an area of interest, in block 40, need not be done with only one menu. The system can be programmed to display one or more further menus, in response to the user's previous selection. However, in this embodiment, the system always selects two or more databases to be searched, and the databases searched are a function solely of the user's responses to the menu(s).

The selection made in block 40 can also be done without conventional menus. Any other means by which the user can indicate, to the system, an area of interest is included within the scope of block 40 of Figure 2, and should be deemed an equivalent.

Another embodiment of the present invention provides an additional

level of standardization for database searching. In this embodiment, the user employs one standardized set of commands for conducting searches in disparate databases. Before describing this embodiment, it is helpful to explain the concept of a "command set" for database searching.

As explained above, each commercial database, or family of databases, has its own rules of syntax which governs the construction of search requests. Each set of rules of syntax includes a set of logical operators used to connect the terms of a search request. Examples of rules of syntax for various databases or database families are given in U.S. Patent No. 4,774,655.

It is not enough to learn the search syntax for a particular database or database family. It is also necessary to know a set of commands needed for conducting a search. Such commands are used for such functions as instructing the system to begin a search, displaying the results of a search, providing information on the types of searching available, terminating a research session, and more. In general, each database, or database family, uses a different set of commands.

In the present invention, the user transmits commands selected from a standardized set. The system translates the search request into the equivalent command applicable to the database or database family being searched.

Figure 4 is a table showing a hypothetical standardized set of commands, with a brief indication of the meaning of each. The table also shows the equivalents of each command in the BRS, DIALOG, and VU/TEXT families of databases. For example, the hypothetical standardized command which directs the system to search the database is "FIND". In DIALOG, the corresponding command is "SELECT". In BRS, it is "..SEARCH". In the hypothetical standardized command set, "SHOW" is used to display

the results of a search. The corresponding BRS command is "..PRINT", the DIALOG command is "TYPE", and the VU/TEXT command is "PL". In the standardized command set, "STOP" ends the searching session; the corresponding command in BRS is "..OFF" and in DIALOG it is "LOGOFF".

The system of the present invention stores, in a suitable memory device, a table of the type shown in Figure 4. When the user enters a command from the standardized set, the system selects the corresponding command from the command set for the database or database family being searched, and transmits that command to the database. The user need not "see" the actual command which is transmitted to the database. It is as if the user is communicating directly with the database using only the standardized commands.

Some of the commands listed in Figure 4 must be used with other parameters. For example, the "FIND" command is followed by a search request. The system translates the word "FIND" into the appropriate command, and may also translate the search request itself into the syntax of the particular database. In another example, the command "SHOW" may be followed by a specification of the documents desired to be displayed. Preferably, the system is programmed to display an error message to the user if the command does not include the required number or type of parameters.

The present invention resides not in the specific commands which form the standardized command set, but in the concept of providing standardization. Any other choice of a standardized command set could be used.

The standardized command set can be combined with other variations of the invention. For example, one way of using the standardized com-

21

mands is in the embodiment of the cited patent wherein the user selects a particular database. The user may know the names and general coverage of a variety of databases, but may not know (and may not want to learn) the command sets appropriate for each database. The system of the present invention can therefore translate standardized commands into the commands recognized by the selected database or database family. The user conducts the search while the system remains connected, or "on-line", with respect to the database. From the viewpoint of the user, it is as if he or she is directly connected to that database. However, the user employs the standardized command set, while the system acts as an intermediary, translating all commands into the commands appropriate to the particular database.

Note that some of the standardized commands (such as "FIND", in the example given above), relate to the process of performing the search, whereas other commands (such as "SHOW", in the above example) relate to the displaying of results. Thus, not all of the standardized commands would be available to the user, in every variation of the invention. For example, if the system selects databases and performs searches entirely automatically, the user would never need to enter a "FIND" command, as the system would do this automatically. But the standardized command set could still be used for directing the display of search results, after the searching is concluded. After the system disconnects from the database and displays the number of documents retrieved, the user can then transmit, for example, a "SHOW" command, followed by the number of the document desired to be displayed. The system would then reconnect to the database, perform the search again, download the requested document, disconnect from the database, and display the document to the user. The above-described arrangement could be used with the embodiment of Figure

2, wherein multiple databases are searched, as well as with the embodiments of the cited patent, which involves searching in a single database at one time. The above-described arrangement could also be done where the system remains connected to the database, and wherein it is not necessary to disconnect and reconnect.

In another alternative, one could provide a system in which the search request syntax is translated but the search commands are not translated. In this embodiment, the user does not know the syntax of a particular database, but does know the command set. This and other similar alternatives are within the scope of the invention.

As stated above, the concept of a database "language" also includes commands relating to field structures. The documents stored in most databases are arranged in segments or "fields". It is usually possible to search a database by fields. Thus, one can request all documents containing the name "Smith" in the "author" field. One way of specifying a field is to use a command known as a "field tag". In the above example, one might enter "AU=SMITH". This command would cause the system to search each author field of each item in the database for the word "SMITH". In general, each database contains a different set of fields, and each database or database family uses a different set of field tags. For example, the "author", "title", and "date" fields could be represented by "AU=", "TI=", and "DA=", respectively. Some database families employ field tags which are suffixes, e.g. "SMITH/AU".

In addition to standardizing of commands for displaying of search results, the invention also includes providing a standardized set of field tags. When the user wants to search by author, or title, or by some other field, the user can enter the search command in a standardized

format (such as "AU=SMITH") and using a standardized set of field tags.
The expression "AU=SMITH" is called a "field command" in this specifica-
tion. The system can then translate the command into the format appro-
priate to the database being searched. Also, the system can be pro-
grammed to translate the operand (e.g. "SMITH") into an format appropri-
ate to a particular database. For example, some databases contain author
information with the last name first, and others place the last name
last. The system could be programmed to "know" that, for certain data-
bases, the standardized command "AU=JOHN SMITH" must be translated into
"AU=SMITH, JOHN".

The invention thus includes at least three possible levels of trans-
lation. The system can translate search requests, commands, and field
commands. Any combination of these types of translations can be incorpo-
rated into a given embodiment of the invention. Thus, for example, one
can provide a system in which the user, not the system, selects a data-
base, wherein the user is expected to know the search syntax of that
database, but wherein the system translates search commands from a stand-
ardized set. This variation is appropriate for sophisticated users who
are familiar with the search syntax of their favorite databases, but who
do not want to memorize different sets of search commands. One can even
provide standardization of field commands only. The invention is not
limited to the above-described combinaticns, however.

One of the problems in database searching is the retrieval of too
few or too many documents, or "postings". As used herein, a "posting"
occurs when a word in the search request is found in a document contained
in the database. The term "document" is used herein in a general sense,
and includes a record retrievable from a database, whether it be an arti-
cle, a patent, or information recorded in any other format. A search

24

which yields no postings is usually of no value to the user, unless the user is trying to verify that a document does not exist. A search which yields a large number of postings, e.g. more than one hundred, is almost as valueless, because it is usually not economical to browse through all the retrieved documents.

In the embodiment to be described, it is assumed that the database (or host) provides a step-by-step display of search postings. That is, suppose that the search request is "DESKTOP AND PUBLISHING". Then the host will display not only the number of documents in the database containing both "DESKTOP" and "PUBLISHING", but will also separately display the number of documents containing "DESKTOP" and the number containing "PUBLISHING". Many of the databases now available provide this kind of intermediate display.

The following embodiment address the problems wherein a search yields either no postings or too many postings. The term "too many" is, of course, a subjective matter, but is defined, for purposes of discussion, as one hundred or more postings.

Searches which yield no postings can be grouped into one of two categories. In the first category, called "Condition A", at least one of the units of the search request is not found in the database. It is assumed that any given search can be represented as one or more "units", or groups of terms, connected by an "AND" operator or its equivalent. A unit can be a word, or a group of words joined by the "OR" operator. For example the search requests "(ENERGY OR CRISIS) AND (OIL OR PETROLEUM)" and "ENERGY AND PETROLEUM" both contain two units. Thus, if the search request includes four words, joined by "AND", (or, alternatively, by a proximity connector specifying that the words must be within a certain

25

number of words of each other), and if, say, the third word is not in the database, the search will yield no postings because of that third word. In the latter example, the system will know that it was the third word which caused the search to fail. The system then displays a message to the user, stating that there were no postings for this word. The system asks the user if the word is spelled correctly or whether it was entered in the proper format. By "format" it is meant that the word may have been restricted to a particular field (e.g. title, author, or abstract), and the field designator may be incorrectly entered. If the spelling or format is incorrect, the user can enter the corrected term, and the search will be executed again.

If the spelling and format are already correct, the system gives the user suggestions for modifying the search. These suggestions can include 1) entering a related term instead of the term that caused the search to fail, 2) relaxing field restrictions (e.g. searching for all occurrences of the word instead of limiting the search to the title, author, or abstract), and 3) deleting the term from the search. The user may also choose to abandon the search at this point. The suggestions are preferably arranged in a menu, and the user can easily make one or more choices.

If there are more than one terms that would cause the search to fail, the system would attempt to resolve each problem separately before the search is resubmitted. That is, the system performs all the operations described above for each search term or unit which yielded no postings, before the search can be performed again.

The second category of searches which yield no postings is called "Condition B". In Condition B, all of the units of the search request yield postings, but the full search request yields no postings. In many

cases, the reason for the null result is that the restrictions on the units of the search request are too strict. Thus, if Condition B occurs, the system first displays to the user the number of postings for each of the intermediate steps of the search, so that the user has the opportunity to make changes.

For example, suppose that the user wants to find articles about leveraged buyouts of large corporations. The search request entered by the user is "LEVERAGED BUYOUT". Now suppose that the system automatically translates this request as "LEVERAGED/TI AND BUYOUT/TI", i.e. the system performs the search for all documents containing the words "LEVERAGED" and "BUYOUT" in the title. The step-by-step results displayed to the user might look as follows:

> 500   LEVERAGED/TI
>
> 600   BUYOUT/TI
>
> 0   LEVERAGED/TI AND BUYOUT/TI

That is, the system retrieved 500 documents with "LEVERAGED" in the title, and 600 documents with "BUYOUT" in the title, but found no documents containing both words in the title.

The user is now given several options, which can be selected from a menu. First, one or more search terms can be modified, replaced, or deleted. Secondly, the field restrictions can be changed or modified. For example, the user may decide not to restrict the search to occurrences of the terms in the title. Thirdly, the user can also relax the combination restrictions. Thus, for example, if the system had chosen to try "LEVERAGED (2W) BUYOUT", meaning that it searched for documents in which "LEVERAGED" occurs within two words of "BUYOUT", the user may want to relax this restriction to a larger "window". The user could even decide to

replace the search with "LEVERAGED AND BUYOUT", which will retrieve all documents containing both words in any location within the document. Finally, the user is also given the opportunity to abandon the search entirely.

The above examples of search syntax are only hypothetical. In practice, the system could use any other means of informing the user about the restrictions that were initially and automatically placed on the search, and can then give the user the chance to relax such restrictions.

Note that in the cases of both Condition A and Condition B, the system identifies the probable points at which the search failed, and offers options for correction.

The system is programmed to display to the user only the options appropriate to a particular search. Thus, for example, if no field restrictions were entered by the system, the user would not be given the option of broadening the field restrictions. If there are no proximity connectors in the search, the user would not be asked to broaden them. Also, if the search request consists of only one term, the option to delete a term would not be presented.

It is quite possible that the user could resolve a Condition A problem and then encounter Condition B. Also, resolution of either Condition A or Condition B could result in the problem of too many postings, to be discussed below. In anticipation of either case, the system should be programmed to place a limit on the number of failed searches that can be performed by one user in one session.

The problem of too many postings is almost as serious as the problem of no postings at all. When the number of postings is greater than 99 (or any other arbitrarily selected number), the system presents a menu of the following choices to the user. First, the user can choose to view

28

the first ten documents, usually arranged in reverse chronological order,
with the most recent items first. Secondly, the user can be given the
opportunity to add terms to the search, thereby narrowing its scope.
Thirdly, the user can narrow the search by limiting one or more search
words to a particular field, or by tightening proximity connectors. For
example, instead of searching for "MONOCLONAL" and "ANTIBODIES", the user
could search for only those items that contain these words in the title.
Or instead of searching for items containing the words "PARTICLE" and
"ACCELERATOR", the user could search for only those items in which these
words are no more than two words apart. The user can be given the option
of viewing the ten most recent items resulting from one term or phrase of
the search request. Finally, the user can also be given the option of
abandoning the search entirely.

Figure 5 is a flow chart which summarizes the embodiment described
above, for assisting the user in the case of no postings or too many
postings. The system executes a search in block 120. If there are no
postings, as determined in test 122, the system determines, in test 124,
whether each search term (or each unit of a search request) generated
postings. If the answer is no, the system displays the number of post-
ings for each unit, in block 126, and displays a menu of choices to the
user, in block 128. The user is given the option of abandoning the
search. Test 130 determines whether the user wants to abandon, and, if
so, the system stops in block 132. If the user wants to modify the
search, the modification is done in block 134, and the search is executed
again.

If all the search terms or units generated postings, the system pro-
ceeds through blocks 136 and 138, and test 140, in similar fashion. The

user may abandon the search, in block 142, or enter a modification, in block 144.

If the search yields a nonzero number of postings, the system checks whether there are too many postings, in test 146. If not, the system displays the results, in block 148. Block 148 can include any other display steps that may be desired. If there are too many postings, then the system displays the number of postings in block 150 and asks the user for a choice, in block 152. The user may abandon the search, through test 154 and block 156, or may modify the search in block 158.

Another embodiment of the invention is useful both in the case of too many postings and in the case of a "successful" search. In this embodiment, the system ranks the retrieved documents in order of apparent relevance. The ranking is done without actually browsing through the documents.

The principle used in ranking retrieved documents can be illustrated with a simple example. Suppose the user wants information on laptop computers. The system searches for documents containing "LAPTOP" and "COMPUTERS" within two words of each other. Suppose that 87 documents are retrieved. Now suppose that the search is narrowed by specifying that both words must appear in the title. Now, the number of documents may be reduced to 23. Suppose further that the search is again narrowed by specifying that the words be within one word of each other. Now, the number of retrieved documents is reduced to 12. It is very likely that these 12 documents are the most relevant of the original 87.

In general, it is apparent that a document that contains a word of the search request in its title is probably more relevant than a document containing the word in the abstract or elsewhere. A document in which two words of a search request are adjacent is probably more relevant than

one in which the words are widely separated. It is therefore possible to
construct a set of criteria which can be used to rank retrieved docu-
ments, based only on the intermediate step information returned by the
database or host.

A hypothetical set of criteria could be as follows. First, a docu-
ment can be ranked according to whether a search term appears in the ti-
tle and descriptor, the title alone, the descriptor alone, or the ab-
stract. The "descriptor" field is a field containing key words from the
document, and is provided in many databases. Thus, if a search term is
found in both the title and the descriptor field of a document, the docu-
ment is considered the most relevant. If the term is found only in the
abstract, the document is considered the least relevant.

Secondly, the relevance of a retrieved document is related to the
number of words of the search request which appear in the same field of
the document. Thus, a document in which all of the words of the search
request appear in the title is more relevant than one in which the words
of the request appear in different portions.

Thirdly, if the search contains a proximity operator (i.e. if the
search is trying to retrieve documents containing two or more words which
are separated by a given number of words), a retrieved document is con-
sidered most relevant if the words are adjacent, and least relevant if
the words are far apart.

An algorithm can therefore be constructed which ranks the documents
retrieved. The simplest such algorithm simply examines the field in
which a search term appears (e.g. title, abstract, etc.), and ranks the
documents as described above. A more complex algorithm takes into ac-
count the other criteria described above. It has been found, in prac-

31

tice, that only a relatively small number of indicators of relevance are necessary. The most useful criterion is to determine the field in which the search term appears. The number of terms appearing in a given field has also been found to be a useful criterion of relevance.

It is important to note that the ranking of documents by relevance is done without actually viewing the documents. On the contrary, the ranking is done only by determining which terms, or combinations of terms, appear in which portions of the documents. Such information can almost always be obtained without examining the document. Even if the database does not display intermediate steps, one can "probe" the data-base by entering a search for a given word, restricted to a given field, and observing the number of documents retrieved.

In general, in order to rank the documents by relevance, it may be necessary for the system to perform a search, for each given word, more than once. Thus, the system might search for occurrences of a given word, and then might repeat the same search, limiting the second search to, say, the "title" field. While the use of such duplicate searching adds somewhat to the cost of the search, the cost of search time is usu-ally relatively small compared to the cost of displaying search results. Thus, for a slightly higher cost of doing the search, one can obtain con-siderably more information about the relevance of the documents re-trieved, and can thereby avoid the more significant costs of displaying documents that turn out to be nonrelevant.

A method of ranking retrieved documents according to assumed rele-vance can be summarized as follows. Suppose that the searcher wants to find information about parallel processing computers. The system per-forms a search for documents containing the words "PARALLEL", "PROCES-SING", and "COMPUTERS". It also performs searches for documents which

32

contain the above-mentioned words in the title of the document. These

extra searches are not "seen" by the user. Suppose that, in the database

being searched, there are 1500 documents containing "PARALLEL", 1700 doc-

uments containing "PROCESSING", 2500 documents containing "COMPUTERS",

and 500 documents containing all three words. Suppose further that, for

the searches which are restricted to the title, the corresponding numbers

are 100, 150, and 300, respectively, and that ten documents contain all

three words in the title. Note that the system has actually done eight

searches, i.e. two searches for each word separately, and two searches

for the combination of the three words. Clearly, the ten documents last

retrieved are likely to be the most relevant. The search results are far

more useful, as a result of the system having performed the extra search-

es, as compared with the results obtained by simply searching for docu-

ments containing all three words. Although more direct search time was

expended, as compared with a simple search for the three words, the

search results are much more useful, and it was not necessary to display

a multiplicity of documents in order to find the relevant ones.

The following is a list of features, one or more of which can be

included in a particular embodiment of the present invention:

    1. Automatic selection of one or more databases

    2. Automatic selection of host for a given database

    3. Automatic generation of syntax to be used in
       conducting a search

    4. Translation of search request into syntax of
       selected database or database family

    5. Automatic generation of commands to be used in
       conducting a search

    6. Translation of standardized commands into format
       appropriate to database or database family

33

7. Automatic generation of field commands to be used
   in conducting a search

8. Translation of standardized field commands into
   format of database or database family

9. Automatic substitution of alternative form of
   field commands

10. Automatic generation of suggestions for reformulation
    of search requests in cases where no documents or too
    many documents were retrieved

11. Automatic generation of list of retrieved documents
    in the order of their presumed relevance

To the above list, one can add alternative arrangements for format-
ting and displaying the results of searches. Moreover, one can program
the system to repeat the same search periodically to determine whether
relevant documents may have been added to the databases under review.

The features of the above list can be present in virtually all pos-
sible combinations. All of these features have been discussed above, or
in the cited patent. Item No. 9 refers to the automatic provision for
variant forms, such as "SMITH, JOHN" instead of "JOHN SMITH".

Thus, the invention should not be deemed limited to the particular
embodiments described above. The arrangement of components in Figure 1
is only exemplary; many other combinations of computers, databases, and
modems can be used, only some of which have been described explicitly.
The embodiments of Figures 2, 3 and 4 can be used together or separately,
and, as explained above, some features of each embodiment can be used
alone. All such variations should be deemed within the spirit and scope
of the following claims.

34

What is claimed is:

1. A method of supplying information to a user from databases, comprising the steps of:

a)   accepting, from the user, a specification of the subject area in which the user desires to search, and a search request, the search request comprising at least one word for which the user desires to search,

b)   associating at least two databases with the specification of subject made by the user,

c)   automatically executing the search request for each of the databases identified in step (b), and

d)   displaying, to the user, the number of items retrieved, from each database, during each execution of the search request.

2. The method of Claim 1, further comprising the steps of:

a)   accepting, from the user, a designation of at least one of the retrieved items, and

b)   retrieving, from the respective databases, each item designated by the user, and displaying each item to the user.

3. The method of Claim 2, wherein the retrieving step comprises the steps of establishing connection with a database, executing the selected search request, and downloading information obtained from the database.

4. The method of Claim 3, wherein the searches in each database are performed sequentially, and wherein each search includes the steps of establishing connection with a database, performing a search in that database, downloading the results of the search, and terminating the connection with the database.

5. The method of Claim 1, wherein the searches of step (c) are per-

35

formed substantially simultaneously.

6. A method of supplying information to a user, from a plurality of databases, comprising the steps of accepting, from the user, a search request, the search request including at least one word for which the user desires to search, determining an area of interest relating to the search request, automatically selecting a set of databases, the set being determined by said area of interest, the databases in said set being related to said area of interest, and automatically executing the search request in each member of said set of databases.

7. The method of Claim 6, wherein the step of determining an area of interest comprises selecting an area of interest from at least one menu.

8. The method of Claim 6, wherein the searches are performed sequentially in each of said databases.

9. The method of Claim 6, further comprising the step of displaying, to the user, the results of the searches performed.

10. The method of Claim 6, wherein the searches are performed substantially simultaneously in each of said databases.

11. A system for supplying information to a user from databases, the user being located at a terminal, the system comprising a computer, the computer being connected through a first modem to the user's terminal, the computer being connected through at least one second modem to an outgoing telecommunications line, the computer being programmed to accept, from the user, a search request, the search request comprising at least one word for which the user desires to search, the computer comprising means for determining an area of interest related to the search request, the computer also comprising means for associating at least two databases with said area of interest, the computer also being programmed

to execute automatically the search request for each of said databases, by establishing communication with said databases through the second modem and the telecommunications line, the computer also being programmed to display, to the user, information retrieved from the databases.

12. The system of Claim 11, wherein there are a plurality of second modems and telecommunications lines, each second modem being connected to the computer and to a telecommunications line, and wherein the computer includes means for connecting to a number of said second modems, the number being equal to the number of databases which are to be searched, wherein the computer is capable of conducting the searches in the databases substantially simultaneously.

13. In a system for retrieving information from a database, the system being programmed to accept a search request and a choice of a database from a user, the system being capable of executing the search request in the selected database and displaying the results of the search in that database, each of the databases being associated with a set of commands, each set of commands being used to direct the operation of the searches and the displaying of results of the searches in a database, wherein the command sets for at least two of the databases are not identical, the improvement wherein the system includes means for accepting, from the user, a command from a standardized set, in conjunction with a search request in a selected database, and means for translating the command from said standardized set into a corresponding command for the selected database.

14. The system of Claim 13, wherein the databases contain information arranged in searchable fields, and wherein the command set includes commands for retrieving information from specified fields.

15. A method of obtaining information from any one of a plurality of databases, each database having a command set for conducting the operation of searches in the database and for displaying the results of such searches, the command sets for at least two databases being different, the method comprising the steps of:

a) specifying a database, a search request and at least one command, the search request including at least one word desired to be searched in the database, the command being selected from a set of standardized commands,

b) translating the command selected from the standardized set into a command which can be recognized by the selected database, and

c) executing the search request in the selected database, according to the translated command.

16. The method of Claim 15, wherein the databases contain information arranged in searchable fields, and wherein the standardized command set includes commands for retrieving information from specified fields, and wherein the translating step includes translating a standardized command relating to information in one of said fields into a command which can be recognized by the selected database.

17. In a system for retrieval of information from a database, the database being located on at least two different database families, the system including a computer, a first modem connected to accept commands from a user located at a terminal, and a second modem connected to an outgoing telecommunications line, the computer being capable of establishing connection with a plurality of database families, using the second modem and the telecommunications line, the computer being programmed to accept a search request from the user, the search request comprising at least one word for which the user desires to search, the computer com-

prising means for selecting a database, the computer being programmed to transmit the search request to a selected database, the improvement wherein the computer is programmed to determine the database family to be used in gaining access to the selected database, and to gain access to said database through the selected database family.

18. The improvement of Claim 17, wherein the computer is programmed to gain access to an alternative database family which contains said database, if the first-selected database family is unavailable.

19. A system for searching multiple databases comprising a computer, means for connecting the computer to a user so that the user can direct a search request to the computer, a plurality of modems, each modem being connected to an outgoing telecommunications line, each modem being connected to the computer, wherein the computer is programmed to determine a number of databases in which to execute the search request, and wherein the computer is programmed to operate said number of modems to execute the search request in each of said databases substantially simultaneously, wherein the number of modems operated by the computer in response to the search request is not greater than the number of databases in which the search request is to be executed.

20. The system of Claim 19, wherein there are a plurality of users, and wherein the computer is programmed to execute searches substantially simultaneously for said users, and wherein the computer is programmed to take command of the number of modems necessary to execute a search for any one of said users, the computer being programmed to take command of a given modem only when said modem is not being operated in a search for another user.

21. A method of retrieving information from a database, the infor-

mation being supplied to a user, the method comprising the steps of:

a)   transmitting a search request to the database, the search request comprising at least one word for which the user desires to search,

b)   determining the number of documents retrieved from the database which fulfill the search request,

c)   determining whether the number of documents retrieved is zero or whether it is greater than a predetermined value, and

d)   displaying to the user, if the number of retrieved documents is zero or greater than said predetermined value, the numbers of documents retrieved by at least one word of the search request, and suggesting to the user that the search can be modified and resubmitted.

22.   The method of Claim 21, wherein step (d) comprises the step of identifying whether any of the words of the search request were not found in any document of the database, and alerting the user if such condition is true.

23.   The method of Claim 21, wherein the suggesting step is selected from the group consisting of suggesting that the user enter a related word instead of a word of the search request, suggesting that the user correct the spelling of a word of the search request, suggesting that the user delete a word from the search request, and suggesting that the user submit a search which seeks documents in which a word of the search request appears in any portion of a document.

24.   In a system for retrieving information from at least one database, the system including a computer which is programmed to accept a search request from a user, the search request including at least one word for which the user desires to search, the computer being programmed to retrieve documents which fulfill the user's search request, the im-

provement wherein the computer is programmed to determine whether the number of documents retrieved is zero or whether it is greater than a predetermined value, and wherein the computer is programmed to display to the user, if the number of retrieved documents is zero or greater than said predetermined value, the numbers of documents retrieved by at least one word of the search request, and to suggest to the user at least one way in which the search can be modified.

25. The improvement of Claim 24, wherein the system is programmed to display the number of documents containing each word of the search request, and wherein the system is programmed to prompt the user to broaden or narrow portions of the search request according to whether the search retrieved zero documents or too many documents.

26. A method of retrieving information from a database, the information being supplied to a user, the method comprising the steps of:

a) transmitting a search request to the database, the search request comprising at least one word for which the user desires to search,

b) determining the number of documents retrieved from the database which fulfill the search request,

c) ranking the documents retrieved in order of relevance, the ranking step being performed without examining the documents but only by determining the number of documents containing various words of the search request, and
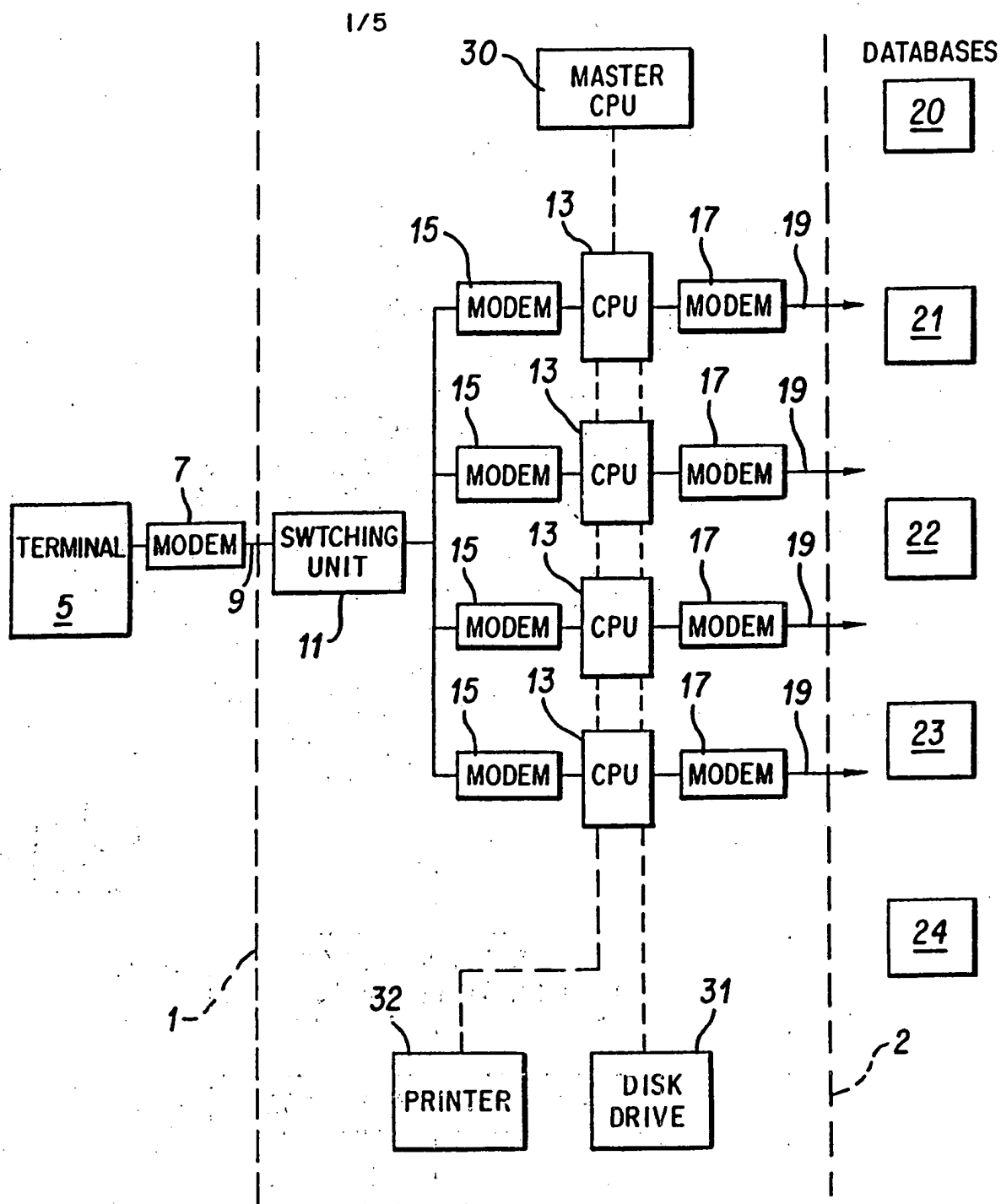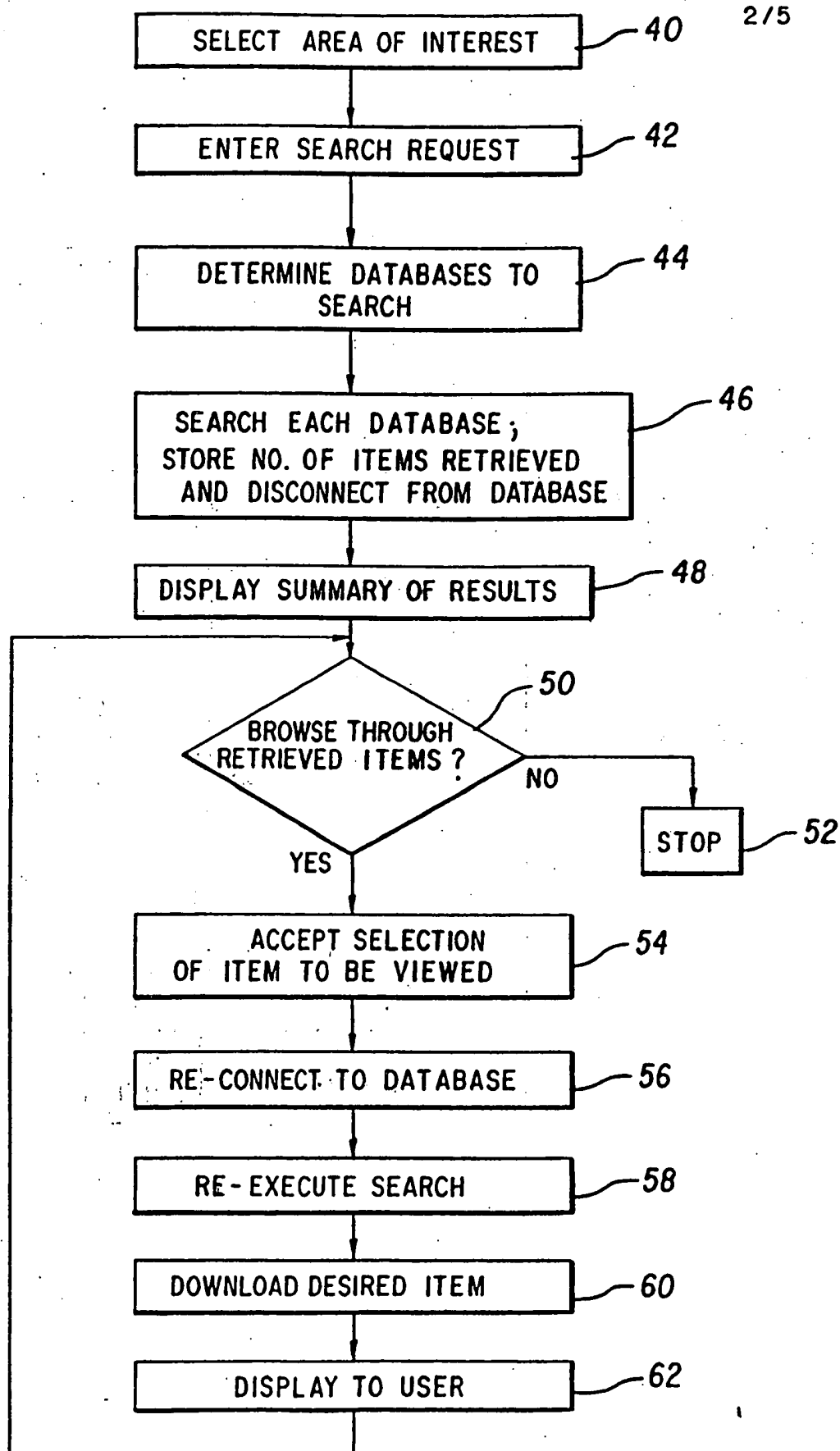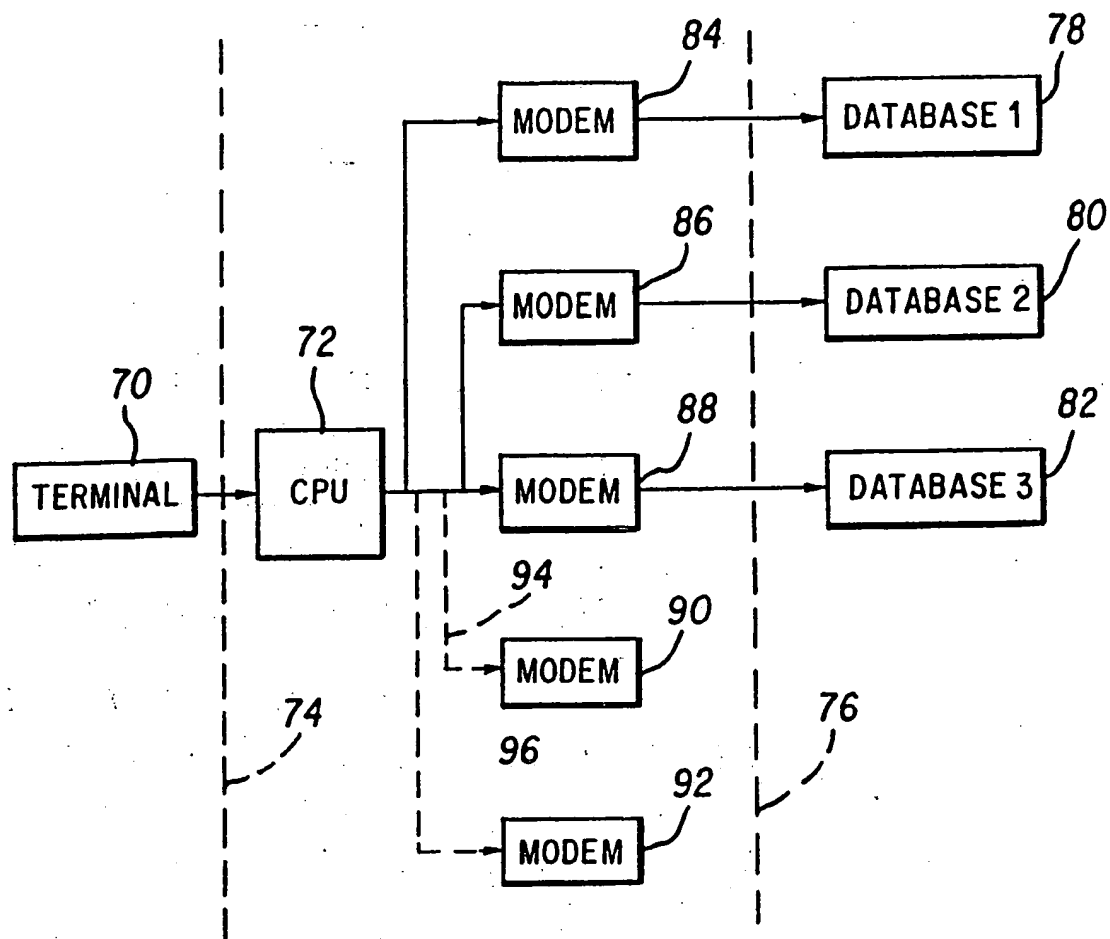
d) displaying, to the user, a predetermined number of the most relevant documents.

27. The method of Claim 26, wherein the documents are divided into a plurality of fields, and wherein the ranking step includes the step of

determining the number of documents containing a word of the search request in a given field of the documents, and ranking the documents in order of the presumed importance of each field.

28. The method of Claim 27, wherein the ranking step includes the step of determining the number of words of the search request which appear in the same field of a document, and ranking the documents in order of said number of words.

29. The method of Claim 26, wherein the ranking step includes the step of determining the number of documents which contain combinations of words of the search request, separated by various numbers of words, and ranking the documents such that the most relevant documents are those in which the words of the search request appear closest together.

30. In a system for retrieving information from at least one database, the system including a computer which is programmed to accept a search request from a user, the search request including at least one word for which the user desires to search, the computer being programmed to retrieve documents which fulfill the user's search request, the improvement wherein the computer is programmed to rank the retrieved documents in order of relevance, without examining the documents but only by determining the number of documents containing various words of the search request.

1/5

DATABASES

FIG. 1

```
┌─────────────────────────────────┐
│     SELECT  AREA  OF  INTEREST  │───── 40
└─────────────────────────────────┘
                 │
                 ▼
┌─────────────────────────────────┐
│     ENTER  SEARCH  REQUEST      │───── 42
└─────────────────────────────────┘
                 │
                 ▼
┌─────────────────────────────────┐
│     DETERMINE  DATABASES  TO    │───── 44
│              SEARCH             │
└─────────────────────────────────┘
                 │
                 ▼
┌─────────────────────────────────┐
│     SEARCH  EACH  DATABASE ;    │───── 46
│  STORE  NO. OF  ITEMS  RETRIEVED│
│  AND  DISCONNECT  FROM  DATABASE│
└─────────────────────────────────┘
                 │
                 ▼
┌─────────────────────────────────┐
│   DISPLAY  SUMMARY  OF  RESULTS │───── 48
└─────────────────────────────────┘
                 │
                 ▼
             ◇ BROWSE THROUGH ◇
             ◇ RETRIEVED ITEMS ? ◇────── 50
                                    NO ──────► ┌────────┐
                 │ YES                         │  STOP  │──── 52
                 ▼                             └────────┘
┌─────────────────────────────────┐
│      ACCEPT  SELECTION          │───── 54
│    OF  ITEM  TO  BE  VIEWED     │
└─────────────────────────────────┘
                 │
                 ▼
┌─────────────────────────────────┐
│    RE-CONNECT  TO  DATABASE     │───── 56
└─────────────────────────────────┘
                 │
                 ▼
┌─────────────────────────────────┐
│      RE - EXECUTE  SEARCH       │───── 58
└─────────────────────────────────┘
                 │
                 ▼
┌─────────────────────────────────┐
│     DOWNLOAD  DESIRED  ITEM     │───── 60
└─────────────────────────────────┘
                 │
                 ▼
┌─────────────────────────────────┐
│       DISPLAY  TO  USER         │───── 62
└─────────────────────────────────┘
```

FIG. 2

FIG. 3

STANDARD COMMAND	MEANING	EQUIVALENTS		
		BRS	DIALOG	VU/TEXT
FIND	SEARCH DATABASE	..SEARCH	SELECT	S
LIST	REVIEW DATABASE DICTIONARY	ROOT	EXPAND	DICT
PHOTO	ORDER PHOTOCOPIES OF ARTICLES	..ORDER	ORDER ITEM	*
REVIEW	REVIEW SETS	..DISPLAY	DISPLAY SETS	RECALL ALL
SHOW	DISPLAY RESULTS	..PRINT	TYPE	PL
STOP	END SESSION	..OFF	LOGOFF	OFF
TOTAL	SHOW SESSION INFORMATION	..TIME	COST	TOT
COMMANDS	DISPLAY COMMANDS	..C	?EXPLAIN	?
INFO	ADDITIONAL HELP	*	?(TOPIC)	HELP

*NOT KNOWN OR NOT AVAILABLE.

FIG. 4
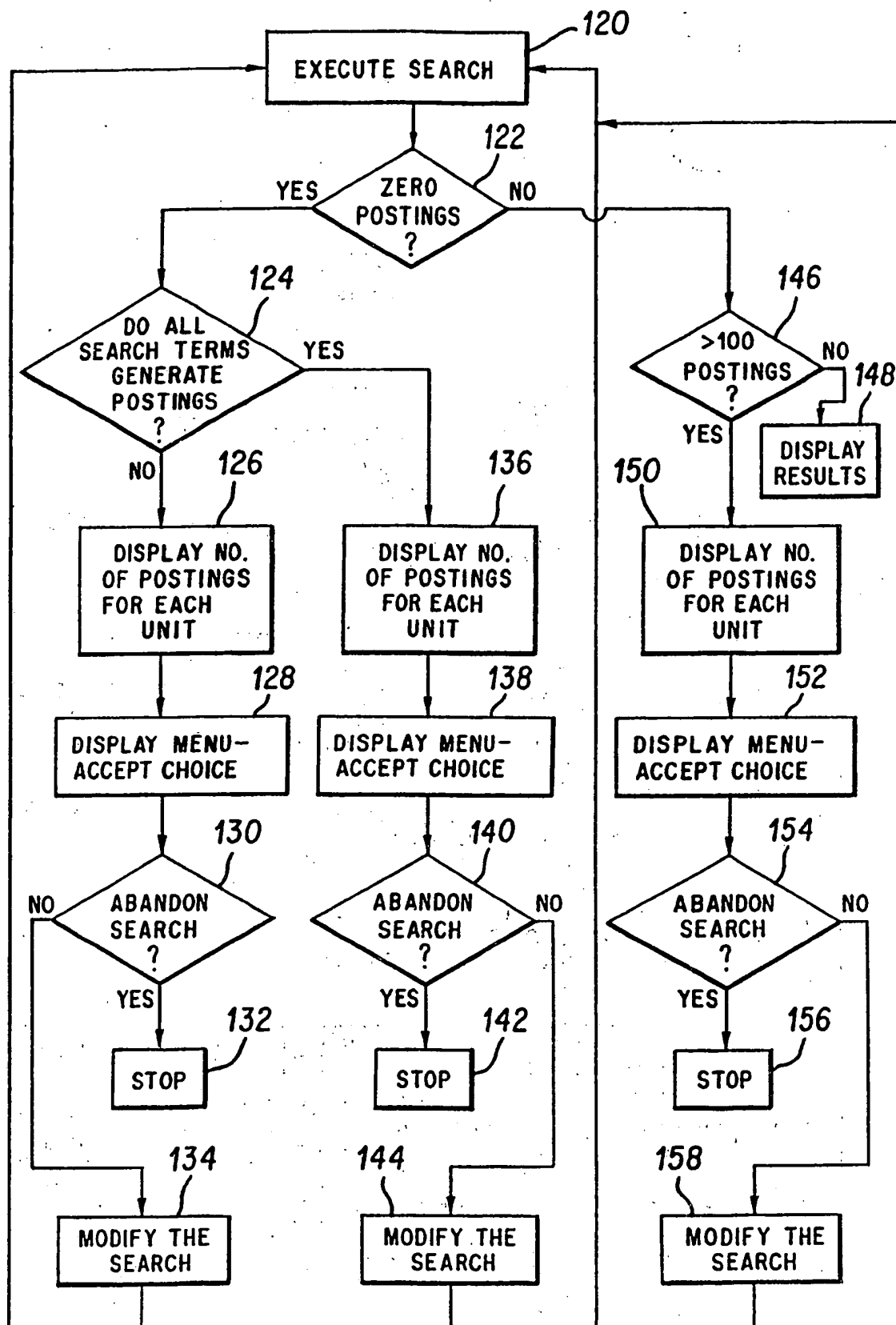
FIG. 5

INTERNATIONAL SEARCH REPORT

I. CLASSIFICATION OF SUBJECT MATTER (if several classification symbols apply, indicate all) *

According to International Patent Classification (IPC) or to both National Classification and IPC

IPC(5): G06F 15/403

U.S. CL.: 364/282.4

II. FIELDS SEARCHED

Minimum Documentation Searched [7]	
Classification System	Classification Symbols
IPC (5):	G06F 15/21, 15/40, 15/403, 15/409.
U.S. CL.:	364/200 364/282.4, 364/900, 364/974

Documentation Searched other than Minimum Documentation
to the Extent that such Documents are Included in the Fields Searched [8]

III. DOCUMENTS CONSIDERED TO BE RELEVANT [9]

Category *	Citation of Document, [11] with indication, where appropriate, of the relevant passages [12]	Relevant to Claim No. [13]
Y	Searching DIALOG The Complete Guide, "Chapter 9 Searching Multiple Files" January 1988-- See pp. (9-1), (9-4), (9-6), (9-14), (9-15) (9-24), (9-25)	1,2,5,6,8-11 13,18,22
Y	Searching DIALOG The Complete Guide, "chapter 1 Introduction to Dialog" August 1987 See pages (1-2) and (1-5)	3,4,20
Y	US, A, 4,774,655 Kollin et al. 27 September 1988 (27.09.88) See figures 1,3a,3b, abstract, Col. 2 (line 45-et seq.), Col, 4 (line 47-et seq.)	7,11,12-17 19
Y,P	US, A, 4,829,423 Tennant et al. 09 May 1989 (09.05.89) See figures 1-16, Col. 3 (line 14-et seq.)	21,23-26
Y	Searching DIALOG The Complete Guide "chapter 8 DIALOG Commands" See pages (8-Sort 1) to (8-Sort6)	26-30

* Special categories of cited documents: [10]

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

IV. CERTIFICATION

Date of the Actual Completion of the International Search	Date of Mailing of this International Search Report
20 MARCH 1990 (20.03.1990)	27 APR 1990
International Searching Authority	Signature of Authorized Officer
ISA/US	ROBERT B. HARRELL

Form PCT/ISA/210 (second sheet) (Rev.11-87)

Category*	Citation of Document. with indication, where appropriate, of the relevant passages	Relevant to Claim No
A	US, A, 4,769,772 Dwyer 06 September 1988 (06.09.88) See Abstract, Figures 1-8, Col. 1 (line 15-et seq.)	1-30
A	US, A, 4,604,686 Reiter et al. 05 August 1986 (05.08.86) See abstract, Figures 1-8, Col. 1 (line 42-et seq.)	1-30

III. DOCUMENTS CONSIDERED TO BE RELEVANT (CONTINUED FROM THE SECOND SHEET)